

Video summary description scheme and method and system of video summary description data generation for efficient overview and browsing

Publication number: CN1382288 (A)

Publication date: 2002-11-27

Inventor(s): GON KIM JAE [KR]; SUNG CHANG HYUN [KR]; MUNCHUR KIM [KR]

Applicant(s): KOREA ELECTRONICS TELECOMM [KR]

Classification:

- international: G06F17/30; G06T9/00; H04N5/76; G06F17/30; G06T9/00; H04N5/76; (IPC1-7): G06T9/00

- European: G06F17/30M5; G06T9/00

Application number: CN20008014746 20000929

Priority number(s): KR19990043712 19991011

Also published as:

CN100485721 (C)

WO0127876 (A1)

KR20010050596 (A)

JP2003511801 (T)

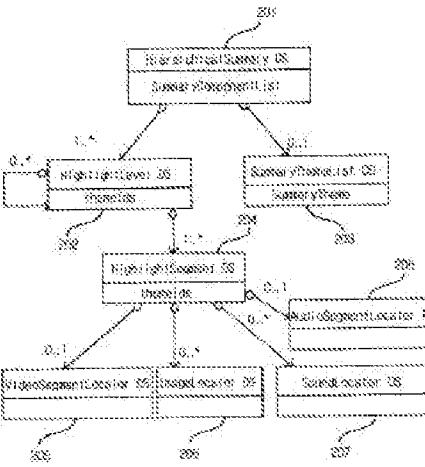
EP1222634 (A1)

[more >>](#)

Abstract not available for CN 1382288 (A)

Abstract of corresponding document: **WO 0127876 (A1)**

The present invention relates to a video summary description scheme for describing video summary by meta data. The video summary provides overview functionality, which makes feasible to understand overall contents of the original video within short time and navigation and browsing functionalities, which make feasible to search the desired video contents efficiently. According to the present invention the HierarchicalSummary DS (DS) comprises at least one HighlightLevel DS and selectively comprises the SummaryThemeList DS. The HighlightLevel DS describe highlight level and may have zero or at least one lower HighlightLevel DS. The HighlightLevel DS comprises one or more HighlightSegment DS which is describing highlight segment information constituting the video summary of the highlight level. The HighlightSegment DS comprises the VideoSegmentLocator DS for describing the time information of corresponding segment interval. Also, the HighlightSegment DS may comprise the ImageLocator DS for describing the representative image information of corresponding segment, the SoundLocator DS for describing the representative sound information, and the AudioSegmentLocator DS for describing the audio segment information constituting the audio summary.



Data supplied from the **esp@cenet** database — Worldwide

[12] 发明专利申请公开说明书

[21] 申请号 00814746.9

[43] 公开日 2002 年 11 月 27 日

[11] 公开号 CN 1382288A

[22] 申请日 2000.9.29 [21] 申请号 00814746.9

[30] 优先权

[32] 1999.10.11 [33] KR [31] 1999/43712

[86] 国际申请 PCT/KR00/01084 2000.9.29

[87] 国际公布 WO01/27876 英 2001.4.19

[85] 进入国家阶段日期 2002.4.23

[71] 申请人 韩国电子通信研究院

地址 韩国大田市

[72] 发明人 金在坤 张现盛 金纹哲 金镇雄

[74] 专利代理机构 北京市柳沈律师事务所

代理人 马莹 邵亚丽

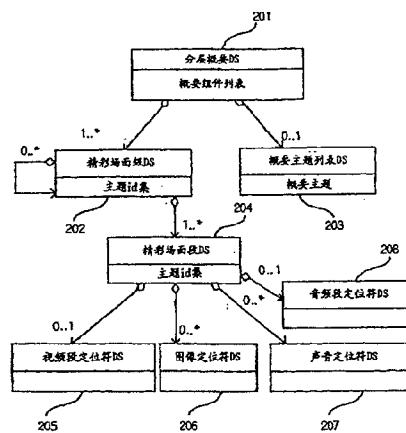
权利要求书 5 页 说明书 13 页 附图 4 页

[54] 发明名称 用于有效概览和浏览的视频概要描述方案和生成视频概要描述数据的方法和系统

[57] 摘要

本发明涉及一种通过元数据描述视频概要的视频概要描述方案。视频概要提供概览功能，这使得在短时间内了解原始视频的全部内容具有可行性，以及导航和浏览功能，这使得有效地搜索所需视频内容具有可行性。根据本发明，分级概要 (HierarchicalSummary) 描述方案 (Description Scheme, DS) 至少包括一个精彩场面级 (HighlightLevel) DS，并且选择性地包括概要主题列表 (SummaryThemeList) DS。精彩场面级 DS 描述精彩场面级，并且可以包含零个或至少一个更低的精彩场面级 DS。精彩场面级 DS 包括一个或多个用于描述组成该精彩场面级的视频概要的精彩场面段信息的精彩场面段 (HighlightSegment) DS。精彩场面段 DS 包括用于描述对应段的代表性图象信息的图象定位符 (ImageLocator) DS，用于描述代表性声音信息的声音定位符 (SoundLocator) DS，和用于描述组成音频概要的音频段信息的音频段定位符 (AudioSegmentLocator) DS。

括用于描述对应段的代表性图象信息的图象定位符 (ImageLocator) DS，用于描述代表性声音信息的声音定位符 (SoundLocator) DS，和用于描述组成音频概要的音频段信息的音频段定位符 (AudioSegmentLocator) DS。



1. 一种用于描述视频概要的分级概要描述方案(DS)，该分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，其中，所述精彩场面级 DS 5 至少包含一个描述组成该精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS。
2. 如权利要求 1 所述的分级概要描述方案，其中，所述精彩场面级 DS 包括至少一个更低级的精彩场面级 DS。
3. 如权利要求 1 所述的分级概要描述方案，其中，所述精彩场面段 DS 10 包括一个描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。
4. 如权利要求 3 所述的分级概要描述方案，其中，所述精彩场面段 DS 进一步包括描述所述对应精彩场面段的代表性帧的图象定位符 DS。
5. 如权利要求 3 所述的分级概要描述方案，其中，所述精彩场面段 DS 15 进一步包括描述所述对应精彩场面段的代表性声音信息的声音定位符 DS。
6. 如权利要求 3 所述的分级概要描述方案，其中，所述精彩场面段 DS 进一步包括描述所述对应精彩场面段的代表性帧的图象定位符 DS 和描述所述对应精彩场面段的代表性声音信息的声音定位符 DS。
7. 如权利要求 4 所述的分级概要描述方案，其中，所述图象定位符 DS 20 描述与所述对应精彩场面段对应的视频区间代表性帧的时间信息或图象数据。
8. 如权利要求 3 所述的分级概要描述方案，其中，所述精彩场面段 DS 进一步包括描述组成所述对应精彩场面段的音频概要的音频段信息的音频段定位符 DS。
9. 如权利要求 8 所述的分级概要描述方案，其中，所述音频段定位符 DS 描述所述对应精彩场面段的音频区间的时间信息或音频数据。
10. 如权利要求 1 所述的分级概要描述方案，其中，所述分级概要 DS 包括描述并且列举包含在分级概要 DS 中的所有概要组件类型的概要组件列表。
- 30 11. 如权利要求 10 所述的分级概要描述方案，其中，所述概要组件类型包括：关键帧，表示由代表性帧组成的关键帧概要；关键视频片断，表示

由关键视频段集组成的关键视频片断概要；关键事件，表示对应于事件或主题的视频区间的概要；关键音频片断，表示由代表性音频区间集组成的关键音频片断概要；和无约束，表示除所述概要之外的由用户定义的概要类型。

12. 如权利要求1所述的分级概要描述方案，其中，所述分级概要 DS 5 包括列举包含在概要中的事件或主题，并且描述 ID 的概要主题列表 DS，然后描述基于事件的概要，并且允许用户通过在所述概要主题列表中描述的事件或主题浏览概要视频。

13. 如权利要求11所述的分级概要描述方案，其中，所述概要主题列表 DS 包括任意数目的概要主题作为元素，并且所述概要主题包括表示对应 10 事件或主题的 id 属性。

14. 如权利要求13所述的分级概要描述方案，其中，所述概要主题进一步包括描述上一级的事件或主题 id 的父 ID 属性。

15. 如权利要求13所述的分级概要描述方案，其中，如果组成对应精彩场面级的所有精彩场面段和精彩场面级具有共同的事件或主题，所述精彩场面级 DS 包括描述共同事件或主题所述 id 属性的主题 id 集属性。

16. 如权利要求13所述的分级概要描述方案，其中，所述精彩场面段 DS 包括描述所述 id 属性的主题 id 集属性，并且描述对应精彩场面段的事件或主题。

17. 一种其中存储有分级概要描述方案(DS)的计算机可读记录介质，该 20 分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，其中所述精彩场面级 DS 至少包括一个描述组成那个精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS，其中所述精彩场面段 DS 包括描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。

18. 一种用于通过输入原始视频根据视频概要描述方案生成视频概要描 25 述数据的方法，包括如下步骤：

 视频分析步骤，通过输入原始视频然后分析原始视频，产生视频分析结果；

 概要规则定义步骤，定义用于选择概要视频区间的概要规则；

30 概要视频区间选择步骤，通过输入所述原始视频分析结果和所述概要规则，从原始视频选择能够概括视频内容的视频区间，组成概要视频区间信息；
 和

视频概要描述步骤，通过输入由所述概要视频区间选择步骤输出的概要视频区间信息，根据分级概要 DS 产生视频概要描述数据。

19. 如权利要求 18 所述的视频概要描述数据生成方法，其中，所述分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，其中所述精彩场面级 DS 至少包括一个描述组成那个精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS，其中所述精彩场面段 DS 包括描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。

20. 如权利要求 18 所述的视频概要描述数据生成方法，其中，所述视频分析步骤包括：

10 特征提取步骤，通过输入原始视频并且提取特征，输出这些特征类型和检测到这些特征的视频时间区间；

事件检测步骤，通过输入所述特征类型和检测到这些特征的视频时间区间，检测包含在原始视频中的关键事件；和

15 插曲检测步骤，通过根据所述检测到的事件将原始视频划分为情节流基本单元，检测插曲。

21. 如权利要求 18 所述的视频概要描述数据生成方法，其中，所述概要规则定义步骤在将概要事件类型定义给所述视频概要描述步骤之后，提供作为选择概要视频区间基础的概要事件类型。

22. 如权利要求 18 所述的视频概要描述数据生成方法，该方法进一步包括代表性帧提取步骤，通过输入所述概要视频区间信息并且提取代表性帧，将该代表性帧提供给所述视频概要描述步骤。

23. 如权利要求 18 所述的视频概要描述数据生成方法，该方法进一步包括代表性声音提取步骤，通过输入所述概要视频区间信息并且提取代表性声音，将该代表性声音提供给所述视频概要描述步骤。

25 24. 一种其中存储有一个程序的计算机可读记录介质，该程序执行如下步骤：

特征提取步骤，输出特征类型和检测到这些特征的视频时间区间；

事件检测步骤，通过输入所述特征类型和检测到这些特征的所述视频时间区间，检测包含在原始视频中的关键事件；

30 插曲检测步骤，通过根据所述检测到的关键事件将原始视频划分为情节流基本单元，检测插曲；

概要规则定义步骤，定义用于选择概要视频区间的概要规则；

概要视频区间选择步骤，通过输入所述检测到的插曲和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和

5 视频概要描述步骤，通过输入由所述概要视频区间选择步骤输出的概要视频区间信息，使用分级概要 DS 生成视频概要描述数据。

25. 一种通过输入原始视频根据视频概要描述方案生成视频概要描述数据的系统，包括：

视频分析装置，用于通过输入原始视频并且分析原始视频，输出视频分析结果；

10 概要规则定义装置，用于定义用来选择概要视频区间的概要规则；

概要视频区间选择装置，用于通过输入所述原始视频分析结果和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和

15 视频概要描述装置，用于通过输入由所述概要视频区间选择装置输出的概要视频区间信息，使用分级概要 DS 生成视频概要描述数据。

26. 如权利要求 25 所述的视频概要描述数据生成系统，其中，所述分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，其中所述精彩场面级 DS 至少包括一个描述组成那个精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS，其中所述精彩场面段 DS 包括描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。

27. 如权利要求 25 所述的视频概要描述数据生成系统，其中，所述视频分析装置包括：

特征提取装置，用于通过输入原始视频并且提取特征，输出这些特征类型和检测到这些特征的视频时间区间；

25 事件检测装置，用于通过输入所述特征类型和检测到这些特征的视频时间区间，检测包含在原始视频中的关键事件；和

插曲检测装置，用于通过根据所述检测到的事件将原始视频划分为情节流基本单元，检测插曲。

28. 如权利要求 25 所述的视频概要描述数据生成系统，其中，所述概要规则定义装置在将概要事件类型定义给所述视频概要描述装置之后，提供作为选择概要视频区间基础的概要事件类型。

29. 如权利要求 25 所述的视频概要描述数据生成系统，该系统进一步包括代表性帧提取装置，用于通过输入所述概要视频区间信息并且提取代表性帧，将该代表性帧提供给所述视频概要描述装置。

30. 如权利要求 25 所述的视频概要描述数据生成系统，该系统进一步包括代表性声音提取装置，用于通过输入所述概要视频区间信息并且提取代表性声音，将该代表性声音提供给所述视频概要描述装置。

31. 一种其中存储有一个程序的计算机可读记录介质，该程序用于运行如下装置：

特征提取装置，用于输出特征类型和检测到这些特征的视频时间区间；

事件检测装置，用于通过输入所述特征类型和检测到这些特征的所述视频时间区间，检测包含在原始视频中的关键事件；

插曲检测装置，用于通过根据所述检测到的关键事件将原始视频划分为情节流基本单元，检测插曲；

概要规则定义装置，用于定义用来选择概要视频区间的概要规则；

概要视频区间选择装置，用于通过输入所述检测到的插曲和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和

视频概要描述装置，用于通过输入由所述概要视频区间选择装置输出的概要视频区间信息，使用分级概要 DS 生成视频概要描述数据。

32. 一种服务器/客户机环境下的视频浏览系统，包括：

服务器，装备有视频概要描述数据生成系统，该系统通过输入原始视频，根据分级概要 DS，生成视频概要描述数据，并且链接所述原始视频和视频概要描述数据；和

客户机，通过使用所述视频概要描述数据概览所述原始视频并且访问所述服务器的原始视频，对视频进行浏览和导航。

用于有效概览和浏览的视频概要描述方案和
生成视频概要描述数据的方法和系统

5

技术领域

本发明涉及一种用于有效概览和浏览视频的视频概要描述方案，并且还涉及一种生成视频概要描述的方法和系统，以根据视频概要描述方案描述视频概要。

10 本发明所涉及的技术领域为基于内容的视频索引和浏览/搜索，并且将视频概括为所基的内容，然后对它进行描述。

发明背景

15 概括视频的格式主要分为动态概要和静态概要。本发明的视频描述方案用于有效地将动态概要和静态概要描述为统一的描述方案。

一般来说，由于现有视频概要和描述方案简单地提供包含在视频概要中的视频区间信息，现有视频概要和描述方案受限于通过播放概要视频，传达全部视频内容。

20 然而，在很多情况下，需要通过概览全部内容标识和重新访问有关部分的浏览，而不仅仅是通过概要视频概览全部内容。

另外，现有视频概要仅仅提供根据由视频概要提供者确定的标准认为是重要的视频区间。因此，如果用户和视频提供者的标准相互不同，或者用户具有特殊标准，用户就不能获得他们所需的视频概要。

25 也就是说，虽然现有概要视频通过提供若干级的概要视频，允许用户选择所需级的概要视频，但是它使用户的选择程度受限于用户不能通过概要视频内容来选择。

30 标题为“Method and apparatus for video browsing based on content and structure(用于基于内容和结构的视频浏览的方法和装置)”的美国专利5,821,945以压缩的形式表示视频，并且通过该表示提供访问具有所需内容的视频的浏览功能。

然而，该专利采用基于代表性帧的静态概要，并且虽然通过使用视频镜

头的代表性帧，概括出现有静态概要，但是该专利的代表性帧只提供代表镜头的视觉信息，该专利对于使用概要传达信息具有限制。

与该专利相比较，该视频描述方案和浏览方法使用基于视频段的动态概要。

5 ISO /IEC JTC1/SC29/WG11 MPEG-7 输出文档号 N2844 在 1999 年 7 月所公布的 MPEG-7 描述方案(V 0.5)提出视频概要描述方案。由于该方案描述动态概要视频的每个视频段的区间信息，因此，尽管提供描述动态概要的基本功能，但是该方案具有如下方面的问题。

10 首先，一个缺点是它不能提供从组成概要视频的视频段对原始视频的访问。也就是，用户想要根据概要内容和通过概要视频的概览，访问原始视频，以了解更详细的信息。然而现有方案不能满足这一需要。

其次，现有方案不能提供足够的音频概要描述功能。

最后，一个缺点是在表示基于事件的概要的情况下，重复描述和搜索复杂性不可避免。

15

发明概要

本发明的一个目的是提供分级视频概要描述方案，它在包含在概要视频中的每个视频区间，包括代表性帧信息和代表性声音信息，并且使基于用户可定制事件的概要提供用户对概要视频内容的选择，和有效浏览具有可行性，和一种使用描述方案的视频概要描述数据生成方法和系统。

为了实现该目的，根据本发明可执行示例的分级概要(HierarchicalSummary)DS 至少包括一个描述精彩场面级的精彩场面级(HighlightLevel) DS，并且精彩场面级 DS 至少包含描述组成该精彩场面级的概要视频的精彩场面段信息的精彩场面段(HighlightSegment) DS。

25 最好，精彩场面级 DS 包括至少一个更低级的精彩场面级 DS。

更好，精彩场面段 DS 包括一个描述所述对应精彩场面段的时间信息或视频本身的视频段定位符(VideoSegmentLocator) DS。

最好，精彩场面段 DS 进一步包括描述所述对应精彩场面段的代表性帧的图象定位符(ImageLocator) DS。

30 更好，精彩场面段 DS 进一步包括描述所述对应精彩场面段的代表性声音信息的声音定位符(SoundLocator)DS。

最好，精彩场面段 DS 进一步包括描述所述对应精彩场面段的代表性帧的图象定位符 DS 和描述所述对应精彩场面段的代表性声音信息的声音定位符 DS。

更好，图象定位符 DS 描述与所述对应精彩场面段对应的视频区间 的代表性的帧的时间信息或图象数据。
5 表性的帧的时间信息或图象数据。

最好，精彩场面段 DS 进一步包括描述组成所述对应精彩场面段的音频概要的音频段信息的音频段定位符(AudioSegmentLocator)DS。

更好，音频段定位符 DS 描述所述对应精彩场面段的音频区间的时间信息或音频数据。

10 最好，分级概要 DS 包括描述并且列举包含在分级概要 DS 中的所有概要组件类型(SummaryComponentType)的概要组件列表(SummaryComponentList)。

另外，最好，分级概要 DS 包括列举包含在概要中的事件或主题，并且描述 ID 的概要主题列表(SummaryThemeList) DS，然后描述基于事件的概要，并且允许用户通过在所述概要主题列表中描述的事件或主题浏览概要视
15 频。

更好，概要主题列表 DS 包括任意数目的概要主题(SummaryTheme)作为元素，并且所述概要主题包括表示对应事件或主题的 id 属性，并且概要主题进一步包括描述上一级的事件或主题 id 的父 ID(parentID)属性。

20 最好，如果组成对应精彩场面级的所有精彩场面段和精彩场面级具有共同的事件或主题，精彩场面级 DS 包括描述共同事件或主题所述 id 属性的主题 id 集(themeIds)属性。

更好，精彩场面段 DS 包括描述所述 id 属性的主题 id 集(themeIds)属性，并且描述对应精彩场面段的事件或主题。

25 另外，根据本发明，提供一种其中存储有分级概要 DS 的计算机可读记录介质。最好，分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，并且精彩场面级 DS 至少包括一个描述组成那个精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS，并且精彩场面段 DS 包括描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。

30 另外，根据本发明，提供一种用于通过输入原始视频根据视频概要描述方案生成视频概要描述数据的方法。该方法包括如下步骤：视频分析步骤，

通过输入原始视频然后分析原始视频，产生视频分析结果；概要规则定义步骤，定义用于选择概要视频区间的概要规则；概要视频区间选择步骤，通过输入所述原始视频分析结果和所述概要规则，从原始视频选择能够概括视频内容的视频区间，组成概要视频区间信息；和视频概要描述步骤，通过输入由所述概要视频区间选择步骤输出的概要视频区间信息，根据分级概要 DS 5 产生视频概要描述数据。

最好，视频分析步骤包括：特征提取步骤，通过输入原始视频并且提取特征，输出这些特征类型和检测到这些特征的视频时间区间；事件检测步骤，通过输入所述特征类型和检测到这些特征的视频时间区间，检测包含在原始 10 视频中的关键事件；和插曲(episode)检测步骤，通过根据所述检测到的事件将原始视频划分为情节流基本单元，检测插曲。

最好，概要规则定义步骤在将概要事件类型定义给所述视频概要描述步骤之后，提供作为选择概要视频区间基础的概要事件类型。

更好，该方法进一步包括代表性帧提取步骤，通过输入所述概要视频区间信息并且提取代表性帧，将该代表性帧提供给所述视频概要描述步骤。 15

更好，该方法进一步包括代表性声音提取步骤，通过输入所述概要视频区间信息并且提取代表性声音，将该代表性声音提供给所述视频概要描述步骤。

另外，根据本发明，提供一种其中存储有一个程序的计算机可读记录介质。该程序执行如下步骤：特征提取步骤，输出特征类型和检测到这些特征的视频时间区间；事件检测步骤，通过输入所述特征类型和检测到这些特征的所述视频时间区间，检测包含在原始视频中的关键事件；插曲检测步骤，通过根据所述检测到的关键事件将原始视频划分为情节流基本单元，检测插曲；概要规则定义步骤，定义用于选择概要视频区间的概要规则；概要视频 20 区间选择步骤，通过输入所述检测到的插曲和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和视频概要描述步骤，通过输入由所述概要视频区间选择步骤输出的概要视频区间信息，使用分级概要 DS 25 生成视频概要描述数据。

另外，根据本发明，提供一种通过输入原始视频根据视频概要描述方案生成视频概要描述数据的系统。该系统包括：视频分析装置，用于通过输入原始视频并且分析原始视频，输出视频分析结果；概要规则定义装置，用于 30

定义用来选择概要视频区间的概要规则；概要视频区间选择装置，用于通过输入所述原始视频分析结果和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和视频概要描述装置，用于通过输入由所述概要视频区间选择装置输出的概要视频区间信息，使用分级概要
5 DS 生成视频概要描述数据。

最好，分级概要 DS 至少包括一个描述精彩场面级的精彩场面级 DS，精彩场面级 DS 至少包括一个描述组成该精彩场面级的概要视频的精彩场面段信息的精彩场面段 DS，并且精彩场面段 DS 包括描述所述对应精彩场面段的时间信息或视频本身的视频段定位符 DS。

10 最好，视频分析装置包括：特征提取装置，用于通过输入原始视频并且提取特征，输出这些特征类型和检测到这些特征的视频时间区间；事件检测装置，用于通过输入所述特征类型和检测到这些特征的视频时间区间，检测包含在原始视频中的关键事件；和插曲检测装置，用于通过根据所述检测到的事件将原始视频划分为情节流基本单元，检测插曲。

15 更好，概要规则定义装置在将概要事件类型定义给所述视频概要描述装置之后，提供作为选择概要视频区间基础的概要事件类型。

最好，该系统进一步包括代表性帧提取装置，用于通过输入所述概要视频区间信息并且提取代表性帧，将该代表性帧提供给所述视频概要描述装置。

20 更好，该系统进一步包括代表性声音提取装置，用于通过输入所述概要视频区间信息并且提取代表性声音，将该代表性声音提供给所述视频概要描述装置。

另外，根据本发明，提供一种其中存储有一个程序的计算机可读记录介质。该程序用于运行如下装置：特征提取装置，用于输出特征类型和检测到这些特征的视频时间区间；事件检测装置，用于通过输入所述特征类型和检测到这些特征的所述视频时间区间，检测包含在原始视频中的关键事件；插曲检测装置，用于通过根据所述检测到的关键事件将原始视频划分为情节流基本单元，检测插曲；概要规则定义装置，用于定义用来选择概要视频区间的概要规则；概要视频区间选择装置，用于通过输入所述检测到的插曲和所述概要规则，选择能够概括原始视频的视频内容的视频区间，组成概要视频区间信息；和视频概要描述装置，用于通过输入由所述概要视频区间选择装

置输出的概要视频区间信息，使用分级概要 DS 生成视频概要描述数据。

另外，提供一种根据本发明的服务器/客户机环境下的视频浏览系统。该系统包括：服务器，装备有视频概要描述数据生成系统，该系统通过输入原始视频，根据分级概要 DS，生成视频概要描述数据，并且链接所述原始 5 视频和视频概要描述数据；和客户机，通过使用所述视频概要描述数据概览所述原始视频并且访问所述服务器的原始视频，对视频进行浏览和导航。

附图简述

将参照附图对本发明的实施例进行说明，其中：

10 图 1 是示出用于根据本发明的描述方案生成视频概要描述数据的系统的方框图；

图 2 是采用 UML(Unified Modeling Language, 统一建模语言)示出描述本发明的视频概要描述方案的分级概要 DS 的数据结构的图；

15 图 3 是用于播放和浏览输入用与图 2 相同的描述方案描述的视频概要描述数据的概要视频的工具的用户界面组合图；

图 4 是示出使用本发明概要视频的分级浏览的数据和控制流的组合图。

发明的详细描述

将参照附图通过优选实施例对本发明进行详细描述，其中相同的参考号 20 用来标识相同或类似的部分。

图 1 是示出用于根据本发明的描述方案生成视频概要描述数据的系统的方框图。

如图 1 所示，本发明的用于生成视频描述数据的装置包括特征提取部分 101、事件检测部分 102、插曲检测部分 103、概要视频区间选择部分 104、概要规则定义部分 105、代表性帧提取部分 106、代表性声音提取部分 107 25 和视频概要描述部分 108。

特征提取部分 101 通过输入原始特征提取生成概要视频所需的特征。一般特征包括镜头边界、摄像机移动、字幕区域、正面区域等。

在提取特征步骤，通过提取特征，将这些特征类型和检测到这些特征的 30 视频时间区间以(特征类型、特征序列号、时间区间)格式，输出到检测事件步骤。

例如，在摄像机移动的情况下，(摄像机移动，1, 100~150)表示在 100~150 帧中检测到摄像机第一移动的信息。

事件检测部分 102 检测包含在原始视频中的关键事件。由于这些事件必须很好地代表原始视频内容，并且是用于生成概要视频的基准，因此一般根据原始视频种类对这些事件进行不同的定义。
5

这些事件可以表示更高意义层，或者是能够直接推断更高意义的视觉特征。例如，在足球视频的情况下，进球、射门、字幕、回放等可以定义为事件。

事件检测部分 102 以(事件类型，事件序列号，时间区间)输出所检测事件的类型和时间区间。例如，以(射门，1, 200~300)的格式输出表示发生在 10 200 到 300 帧之间的第一射门的事件信息。

插曲检测部分 103，根据所检测的事件，将视频划分为基于情节流的比事件更大单位的插曲。在检测到关键事件之后，检测插曲，同时包括跟随关键事件的伴随事件。例如，在足球视频的情况下，进球和射门可以是关键事件，而教练席场景、观众场景、进球庆祝场景、进球回放场景等组成关键事件的伴随事件。
15

也就是，根据进球和射门检测插曲。

以(插曲号，时间区间，优先级、特征镜头、相关事件信息)格式输出插曲检测信息。在此，插曲号是插曲的序列号，并且时间区间表示以镜头为单位的插曲时间区间。优先级表示插曲的重要度。特征镜头表示包括组成插曲的镜头中最重要信息的镜头号，并且相关事件信息表示与插曲相关的事件的事件号。例如，在将插曲检测信息表示为(插曲 1, 4~6, 1, 5, 进球 1, 字幕 3)的情况下，该信息表示第一插曲包括第 4~6 镜头，优先级为最高(1)，特征镜头为第五镜头，并且相关事件为第一进球和第三字幕。
20

概要视频区间选择部分 104 选择根据所检测的插曲选择能很好概括原始视频内容的视频区间。由概要规则定义部分 105 的预定概要规则执行选择区间的基准。
25

概要规则定义部分 105 定义用于选择概要区间的规则，并且输出用于选择概要区间的控制信号。概要规则定义部分 105 还将用作选择概要视频区间基础的概要事件类型，输出到视频概要描述部分 108。
30

概要视频区间选择部分 104 以帧为单位输出所选概要区间的时间信

息，并且输出与视频区间对应的事件类型。也就是，(100~200，进球)，(500~700，射门)等格式表示选作概要视频区间的视频段为100~200帧，500~700帧等，并且两段的事件分别为进球和射门。另外，可以输出如文件名的信息，来帮助访问仅组成概要视频区间的附加视频。

5 如果完成概要视频区间选择，通过使用概要视频区间信息，分别从代表性帧提取部分106和代表性声音提取部分107提取代表性帧和代表性声音。

代表性帧提取部分106输出代表概要视频区间的图象帧号或输出图象数据。

10 代表性声音提取部分107输出代表概要视频区间的声音数据或输出声音时间区间。

视频概要描述部分108根据图2所示的本发明的分级概要描述方案描述相关信息，以使得有效概览和浏览功能具有可行性。

分级概要描述方案的主要信息包括概要视频的概要事件类型、描述每个概要视频区间的时间信息，代表性帧，代表性声音和每个区间的事件类型。

15 视频概要描述部分108根据图2所示的描述方案输出视频概要描述数据。

图2是采用UML(Unified Modeling Language，统一建模语言)示出本发明的描述视频概要描述方案的分级概要DS的数据结构的图。

20 分级概要DS201描述由一个或多个精彩场面级DS202和一个或零个概要主题列表DS203组成的视频概要。

概要主题列表DS通过列举描述组成概要的主题或事件的信息，提供基于事件的概览和浏览的功能。精彩场面级DS202由若干精彩场面段DS204和零个或若干个精彩场面级DS组成，其中精彩场面段DS204的数目为组成那个级的概要视频的视频区间数。

25 精彩场面段DS描述与每个概要视频区间对应的信息。精彩场面段DS由一个视频段定位符DS205、零个或若干图象定位符DS206、零个或若干声音定位符DS207和音频段定位符208组成。

下面给出关于分级概要DS的更加详细的描述。

30 分级概要DS具有一个概要组件列表属性，该属性清楚地表示由分级概要DS包括的概要类型。

根据概要组件类型得到概要组件列表，并且通过列举所包括的所有概要

组件类型描述该列表。

概要组件列表中存在如关键帧、关键视频片段、关键音频片段、关键事件和无约束这五种类型。

5 关键帧表示由代表性帧组成的关键帧概要。关键视频片断表示由关键视
频区间集组成的关键视频片断概要。关键事件表示由对应于事件或主题的视
频区间组成的概要。关键音频片断表示由代表性音频区间集组成的关键音频
片断概要。并且，无约束表示除所述概要之外的由用户定义的概要类型。

另外，为了描述基于事件的概要，分级概要 DS 可能包括列举包含在概
要中的事件(或主题)并且描述 ID 的概要主题列表 DS。

10 概要主题列表包含任意数目的概要主题作为元素。概要主题具有一个 ID
类型的 id 属性，并且选择性地具有一个父 id 属性。

概要主题列表 DS 允许用户根据在概要主题列表中描述的每个事件或若干主题浏览概要视频。也就是，输入描述数据的应用工具通过分析概要主题
列表 DS 并且将该信息提供给用户，使用户选择所需的主题。

15 此时，在将这些主题列举为简单格式的情况下，如果主题数目很大，可
能就不容易找出用户所需的主题。

因此，通过将主题表示为类似于 ToC(Table of Content, 内容表)的树状
结构，用户可以有效地在找出所需主题之后对各个主题进行浏览。

为此，本发明允许父 id 属性选择性地用在概要主题中。父 id 表示树状
20 结构中的上层元素(上层主题)。

本发明的分级概要 DS 包括多个精彩场面级 DS，并且每个精彩场面级
DS 包括一个或多个对应于组成概要视频的视频段(或区间)的精彩场面段
DS。

精彩场面级 DS 具有 IDREFS 类型的主题 id 集属性。

25 主题 id 集描述共同于对应精彩场面级 DS 的孩子精彩场面级 DS，或包
含在该精彩场面级中的所有精彩场面段 DS 的主题和事件 id，并且该 id 在所
述概要主题列表 DS 中进行描述。

30 主题 id 集可以表示若干事件，并且当进行基于事件的概括时，通过让
主题 id 集表示组成那个级的精彩场面段中共同的主题类型，解决相同 id 不
必要地在组成那个级的所有段中重复这一问题。

精彩场面段 DS 包括一个视频段定位符 DS 和一个或多个图象定位符

DS，零个或一个声音定位符 DS 和零个或一个音频段定位符 DS。

在此，视频段定位符 DS 描述组成概要视频的视频段的时间信息或视频本身。图象定位符 DS 描述视频段的代表性帧的图象数据信息。声音定位符 DS 描述表示对应视频段区间的声音信息。音频段定位符 DS 描述组成音频概要的区间时间信息或音频信息本身。
5

精彩场面段 DS 具有主题 id 集属性。主题 id 集描述，使用定义在概要主题列表中的 id，在所述概要主题列表 DS 中描述的那个主题或事件与对应精彩场面段相关。

主题 id 集可以表示多个事件，并且它是本发明的一个有效技术，通过
10 让一个精彩场面段具有多个主题，解决当对基于事件的概要使用现有方法时，描述每个事件(或主题)的视频段所导致的不可避免的描述重复这一问题。

当描述组成概要视频的精彩场面段时，采用不同于现有分级概要描述方案的方法，只描述精彩场面视频区间的时间信息，为了描述每个精彩场面段的视频区间信息、代表性帧信息、代表性声音信息，通过采用视频段定位符
15 DS、图象段定位符 DS 和声音定位符 DS，本发明通过引入用于描述组成概要视频的精彩场面段 DS，使通过精彩场面段视频的概览和使用段的代表性帧和代表性声音的导航和浏览得以有效的使用。

通过采用能够描述对应于视频区间的代表性声音的声音定位符 DS，在
20 实际情况下通过能够代表视频区间的特征声音，例如，枪响声、喊叫声、足球中的主持人评论(例如，进球和射门)、戏剧中演员姓名、特定词等，通过在短时间内大致了解该区间是否为包含所需内容的重要区间，或者该区间内包含什么内容，进行有效的浏览，而不播放视频区间是可能的。

图 3 是用于播放和浏览输入用与图 2 相同的描述方案描述的视频概要描述数据的概要视频的工具的用户界面组合图。

25 视频播放部分 301 根据用户的控制播放原始视频或概要视频。原始视频代表性帧部分 305 显示原始视频镜头中的代表性帧。也就是，它由一系列尺寸缩小的图象组成。

不采用本发明的分级概要 DS，而采用附加描述方案来描述原始视频镜头的代表性帧，并且可以在随同由本发明的分级概要 DS 描述的概要描述数
30 据一起提供该描述数据时使用。

用户通过单击代表性帧，访问与代表性帧对应的原始视频镜头。

概要视频级 0 代表性帧部分和代表性声音部分 307 和概要视频级 1 代表性帧部分和代表性声音部分 306 分别显示代表概要视频级 0 和概要视频级 1 的每个视频区间的帧和声音信息。也就是说，它由尺寸缩小的代表一系列图象和声音的图标图象组成。

5 如果用户单击概要视频代表性帧部分和代表性声音部分的代表性帧，用户访问对应于代表性帧的原始视频区间。在此，在单击与概要视频的代表性帧对应的代表性声音图标的情况下，播放该视频区间的代表性声音。

概要视频控制部分 302 输入用户选择控制来播放概要视频。在提供多级概要视频的情况下，用户通过级选择部分 303 选择所需级的概要，进行概览和浏览。事件选择部分 304 列举由概要主题列表提供的事件和主题，并且用户通过选择所需事件，进行概览和浏览。总而言之，这实现了用户定制类型的概要。

图 4 是示出使用本发明概要视频的分级浏览的数据和控制流的组合图。

通过使用图 3 的用户界面，采用图 4 的方法访问浏览数据来执行浏览。15 浏览数据是概要视频、概要视频的代表性帧、原始视频 406 和原始视频代表性帧 405。

假定概要视频具有两个级。不用说，概要视频可以具有比两个更多的级。概要视频级 0 401 是以比概要视频级 1 403 更短的时间进行概括的。也就是说，概要视频级 1 比概要视频级 0 包含更多的内容。概要视频级 0 代表性帧 402 是概要视频级 0 的代表性帧，并且概要视频级 1 代表性帧 404 是概要视频级 20 1 的代表性帧。

概要视频和原始视频通过图 3 的视频播放部分 301 进行播放。概要视频级 0 代表性帧在概要视频级 0 代表性帧和代表性声音部分 306 中显示。概要视频级 1 代表性帧在概要视频级 1 代表性帧和代表性声音部分 307 中显示，25 并且原始视频代表性帧在原始视频代表性帧部分 305 中显示。

图 4 所示的分级浏览方法可以具有各种类型的分级路径，如下面示例所示：

情况 1：(1)-(2)

情况 2：(1)-(3)-(5)

30 情况 3：(1)-(3)-(4)-(6)

情况 4：(7)-(5)

情况 5：(7)-(4)-(6)

全面的浏览方案如下所示。

首先，通过观看原始视频的概要视频，了解原始视频的全面内容。在此，概要视频可以播放概要视频级 0 或概要视频级 1。当在观看概要视频之后想要更详细的浏览时，通过概要视频代表性帧标识感兴趣的视频区间。如果正 5 要查找的场景标识在概要视频代表性帧中，通过直接访问代表性帧所连接的原始视频的视频区间，对它进行播放。并且如果需要更详细的信息，用户通过了解下一级的代表性帧，或通过分级了解原始视频代表性帧的内容，可以访问所需的原始视频。

10 虽然这些分级浏览技术对正在播放原始视频时浏览访问所需内容，可能要花很长的时间，但是通过分级代表性帧直接访问原始视频的内容，可以大幅度地减低浏览时间。

现有的一般视频索引和浏览技术以镜头为单位划分原始视频，并且在构成代表每个镜头的代表性帧之后，通过从代表性帧观看所需的镜头，来访问 15 镜头。

在这种情况下，由于原始视频的镜头数很大，在众多代表性帧中浏览所需内容需要花费大量的时间和精力。

在本发明中，通过使用概要视频代表性帧构成分级代表性帧，快速访问所需视频是可行的。

20 情况 1：播放概要视频级 0，并且从概要视频级 0 代表性帧直接访问原始视频。

情况 2：播放概要视频级 0，并且从概要视频级 0 代表性帧选择最感兴趣的代表性帧，并且在与该代表性帧的附近对应的概要视频级 1 代表性帧中 25 标识所需场景，以在访问原始视频之前了解更详细的信息，然后访问原始视频。

情况 3：在情况 2 难以从概要视频级 1 代表性帧访问原始视频的情况下，选择最感兴趣的代表性帧，以获得更详细信息，并且通过邻近该代表性帧的原始视频代表性帧，标识所需场景，然后使用原始帧的代表性帧访问原始视频。

30 情况 4 和 5 是以回放概要视频级 1 开始的情况，路径与上述情况类似。

当应用到服务器/客户机环境时，本发明可以提供其中多个客户机访问

一个服务器，并且可以进行视频概览和浏览的系统。原始视频输入到服务器，根据分级概要描述方案，产生视频概要描述数据，并且装备有链接所述原始视频和视频概要描述数据的概要视频描述数据生成系统。客户机通过通信网络访问服务器，使用视频概要描述数据，对视频进行概览，并且通过访问原
5 始视频，对视频进行浏览和导航。

尽管本发明是根据优先实施例来描述的，但这些实施例对本发明不起限制作用，而只起示例作用。另外，本领域的技术人员应该理解，在不脱离由所附权利要求限定的本发明的精神和范围的情况下，可以对在此的实施例进行修改和变化。

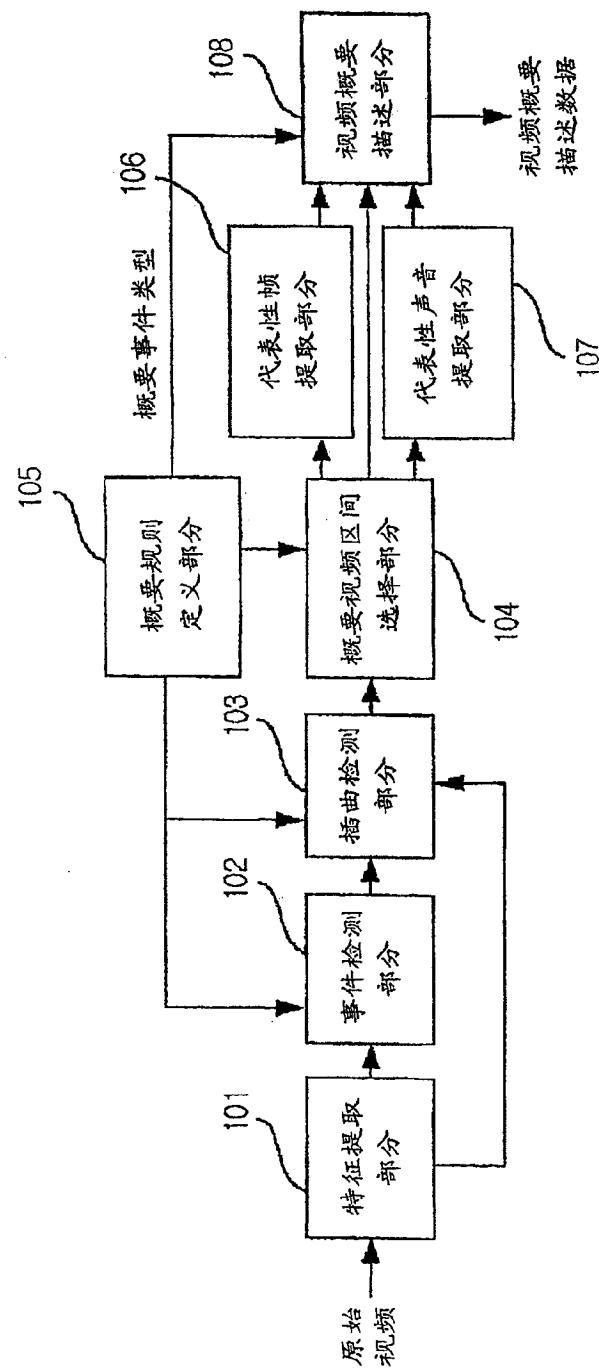


图 1

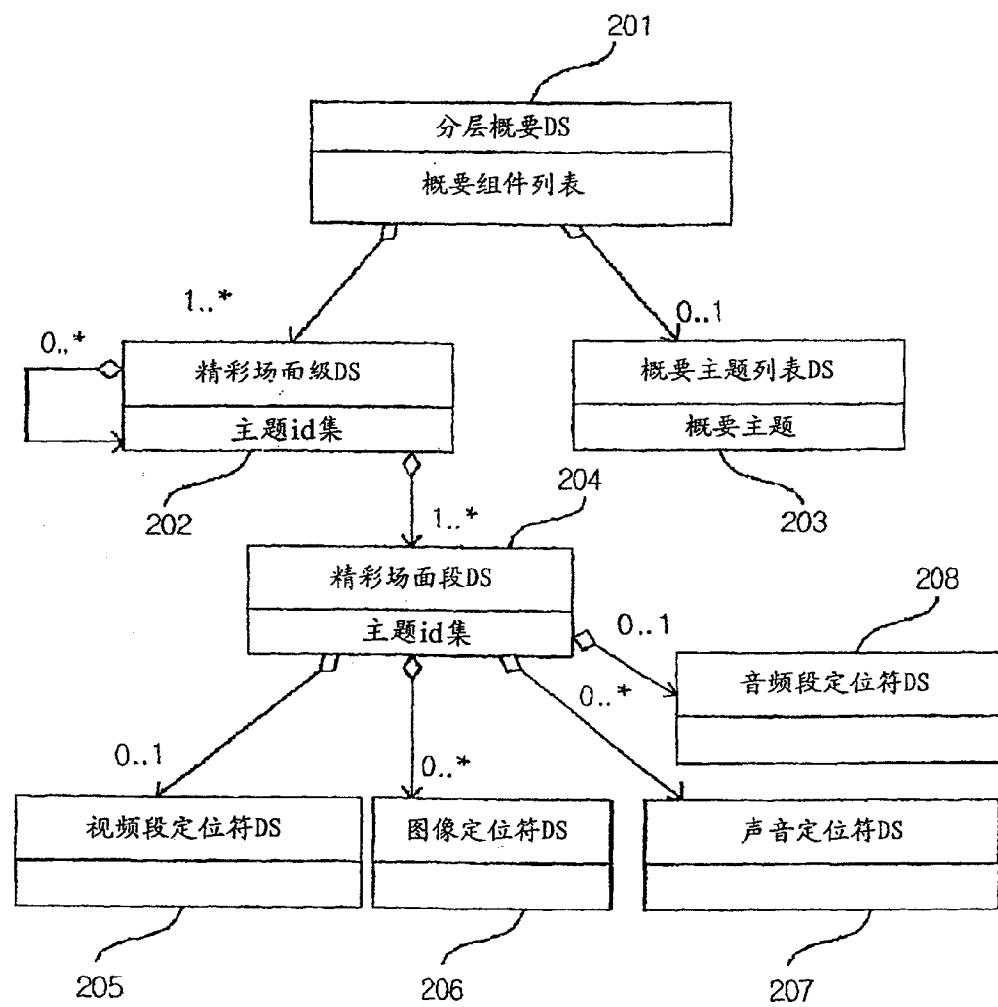


图 2

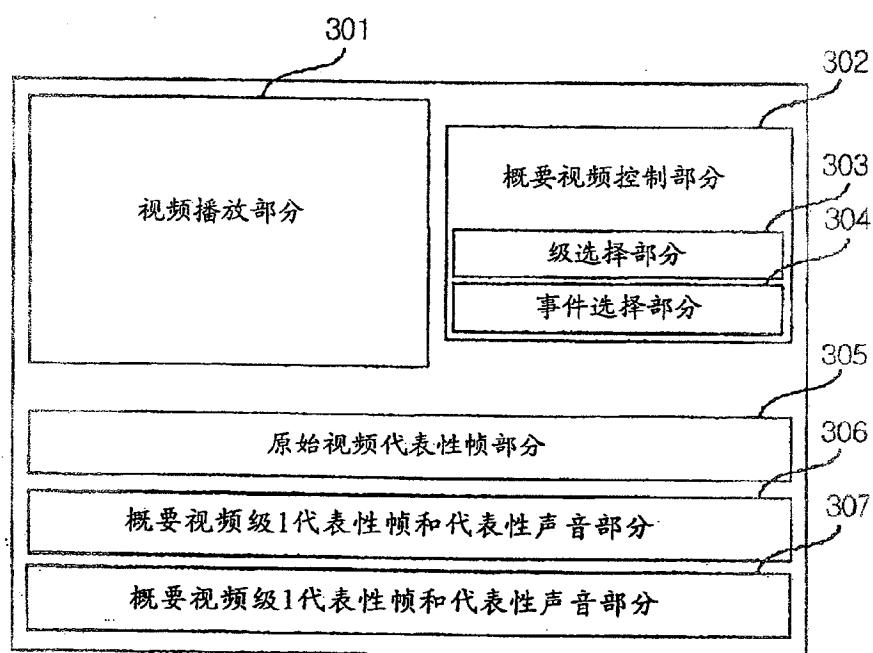


图 3

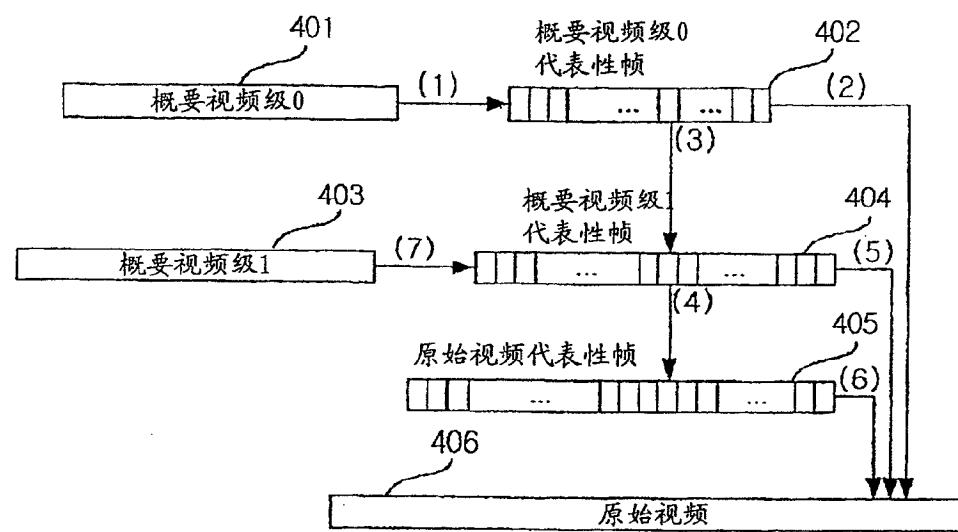


图 4